A guide to TV series success

Alexandre Ferreira
dept. of Computer Science
Faculty of Sciences of the University of Porto
Porto, Portugal
up201906326@edu.fc.up.pt

Mafalda Matos
dept. of Computer Science
Faculty of Sciences of the University of Porto
Porto, Portugal
up201605629@edu.fc.up.pt

Nuno Teixeira
dept. of Computer Science
Faculty of Sciences of the University of Porto
Porto, Portugal
up202204606@edu.fc.up.pt

Abstract—**Success is an arbitrary concept and hard to identify. When together with another unpredictable subject such as film making, the variability created becomes a problem for any meaningful analysis. This paper comes in response to this problem given a clear definition to success based on viewer feedback. Furthermore, the IMDb dataset will serve as a basis for an expansive exploratory analysis with the means to understand and predict what makes a TV series successful. Aspects of film production such as actors, cast and writers, together with a greater examination on genres and episode runtime were considered in order to provide a reliable forecast of a show's performance. Additionally, research on streaming platforms and their influence on the shows was conducted. Finally with this analysis as basis an ideal show capable of achieving success was proposed.**

Keywords— IMDb, Tv series, Success, Streaming Services

I. Introduction

IMDb is the world's most popular and authoritative source for movie, TV and celebrity content. Also knows as the Internet Movie Database, it provides extensive and reliable information about the big and small screen releases alike. This information can range from general knowledge about the titles and crew, to the actual viewer feedback. The report will focus on a subset of IMDb data available for access to customers for personal and non-commercial use.

With the fleeting attention span of today's generation, the need to garner an audience becomes a priority. Attending to this need, an intrinsic need to establish what made titles successful needs to be researched. Whist focussing on Tv Series, this document comes as a response to the problem, trying to understand and predict the predisposition of the global audience and forecasting the necessary steps to make a successful series.

Furthermore, with the growing presence of streaming platforms, an additional public dataset containing information on titles and their showings on the respective platforms will be considered.

II. Dataset

A. IMDb

Before any pre-processing of the data, its essential to acknowledge the subset given and get a greater understanding about the information portrait by it. For starters, the dataset is distributed in 7 different files, each containing unique information. Even though separated, the data is still related to each other using unique identifiers and distinguish with person and title. Because these relationships are hard to visualise, a database model containing each file and their connection with one another was developed, as seen in Figure 1.

As expected from a dataset of this magnitude, not all the information contained in it is up to par. In order to mitigate these problems, the following pre-processing steps took place.
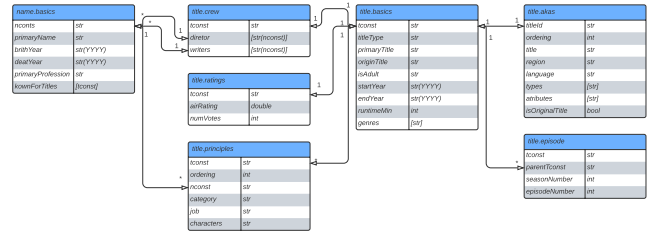

Fig. 1. IMDb Database Model

1) Filtering

The IMDb dataset houses 9916880 different titles, each containing anywhere from 8 to 19 parameters of data. This appended to the 9993719 different people, with at least 5 columns worth of information, make the amount of data in the tables significant, but also computationally intensive. From the established motivation, stated in Chapter I, TV series will be the focus of the analysis allowing for a resampling of the data. This filtering allows for a considerable reduction of the information contained in the dataset while still leaving an extensive amount of data to be analysed, as shown in Figure 2.
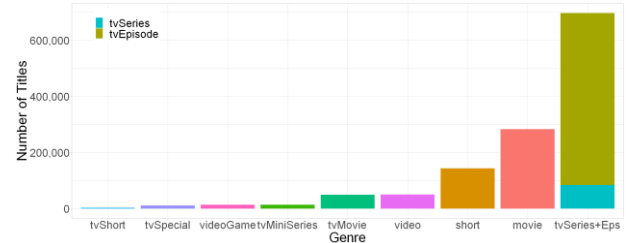

Fig. 2. Number of titles for each type

2) Cleaning

The cleaning efforts started by standardizing the missing values, replacing anything missing and the already existing notation '\NA' and '' by 'NA'. This allows for a straighter forward detection and ensuing treatment. Following that, the missing values were dealt with on a case-to-case basis, establishing different rules for the variables. These rules depend not only on the amount of null data, but also its importance to the overarching problem. For table.basics, the cleaning process was the following:

- Start Year/Genres (13.3%/4.6% NA): The big amount of title data and the low NA percentage, together with the difficult to predict parameters justify ignoring the rows containing the missing values.

- End Year (99% NA): Mostly null, shows that most titles start year and end year are the same, so the column can be discarded.
- Runtime Minutes (73.2%): The high percentage of missing values could justify ignoring the column all together, but even with the limited data a specialized analysis could be done in the future. For that reason, the column will be kept as is.

Important to note, that an additional cleaning/sampling of the data is done with the merging. Specifically, during the merge between title.basics and title.ratings (that has no NA values), about 20% of the titles were removed, because they had no rating. Every plot, including the already displayed Figure 2 represent the number of titles considering this. In addition, some of the column types had to be changed in order to match the expect input given by IMDb.

3) Merging

The model shown on Figure 1, highlights the relationships between tables and the attributes that establish their connections, but the datasets do not represent the same reality. Since the tables are all separated, in their respective file, a merging effort must take place to establish the expected relations. This, even though not mandatory, greatly facilitates the navigation within the data. Because of the size of the dataset, in order to avoid unnecessary waiting times, this process was only made when establishing relations between tables was necessary.

B. Streaming Platform

With the rapid growth of the internet, streaming platforms became a prominent way for people to consume media. Their ease of use and portability, together with affordable prices and exclusive content made them a powerhouse within the entertainment industry. Despite the quality of a series, if they are not available to be watched, or locked in an unreliable steaming platform or cable network it will not be seen.

To enrich the analysis in that regard, the dataset "TV shows on Netflix, Prime Video, Hulu and Disney+" by Ruchi Bhatia on TV series and their respective availability in the main streaming platforms was used. The dataset contains the title of the show together with the IMDb and Rotten Tomatoes ratings as well as the streaming platforms where the show is available - Netflix, Hulu, Prime Video and Disney Plus.

1) Dataset characteristics

The dataset used is much smaller than the datasets previously used, but it is still large enough to where it is statistically relevant, especially given the nature of streaming services (with aprox. 5300 entries). Not only that, but the dataset has been updated as of 2021, which means it should be on par with the information from the previous datasets.

It was also verified that the only show all these platforms had in common was Genius (2017). So, while there is some communality between the streaming services, it is not as large as one might expect.

2) Preprocessing

Unlike the IMDb dataset the pre-processing efforts of the current dataset were more straight forward. For starters, and because of their limited occurrence all rows containing NA information were promptly deleted. Afterwards the only remaining process is to homogenize the ratings, converting

‘x.x/10’ format from IMDb to the desired numeric quantifier used on the previous dataset. Since the platform data was organized as a binary matrix, those numbers were also converted to numeric for ease of use.

3) IMDb vs. Rotten Tomatoes Rating

From a superficial analysis of the dataset, there can be a significant difference between IMDb and Rotten Tomatoes ratings for the same production. For example, Genius (2017) has a rating of 8.3/10 on IMDb, while on Rotten Tomatoes it has a rating about 10% lower at 74/100. This finding puts into question whether IMDb should be used as the sole platform for this study.

III. SUCCESS QUANTIFIERS

Success while globally understood, is still a very ambiguous concept and hard to fully grasp. When it comes to the entertainment industry the most normalized success criteria is how much money the title made in relation to the amount of money used for its production, or the profit. This analysis will not touch on that aspect, since the production costs are not usually publicly available, and the more well-known ones are still just estimations from the media. The gross values are also not straight forward, because when talking about TV series most of the money generated nowadays comes from streaming deals with streaming platforms as well as tv networks, which are most of the times kept private. That said, from the available information, two clear success quantifiers can be defined:

1) Average Rating

The most prevalent and well-known way to establish the quality of the movie, also carries a lot of wight when it comes to the community's opinion on the title. A straightforward ranking system going from 0 (worst) to 10 (best) is usually used in parallel with the gross profit of the series to evaluate its success.

2) Number of votes

Adjacent to average rating and arguably just as important, the number of votes carry two important meanings. Firstly, can be used as an engagement measurement, providing the number of people that after watching a title carried on engaging with it by rating it. Secondly, influences the weight of the average rating by neglecting shows with low number of votes and given more credibility to titles with many voters. This attributes on their own are enough to establish relations between series and draw boundaries of success, as explicit in Figure 3, where the distribution of the average rating and the number of votes for all TV series can be seen.
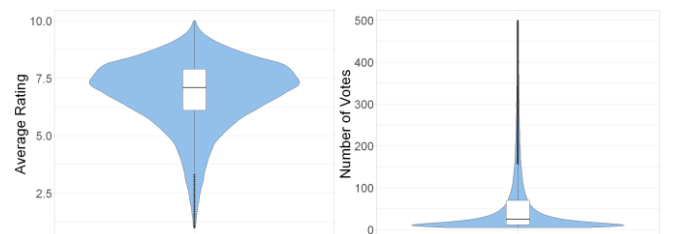

Fig. 3. TV Series Distribution (a) Average Rating, (b) Number of Votes (some outliers removed to better understand how the values are distributed)

In addition to this metrics, the following attribute was evaluated because of their theorical even though less apparent connection to success:

3) Longevity

How long a series lasts can be an interesting indicator of its presence and impact on the community. This was validated by comparing how the average rating and number of votes evolve in relation to the age of the series.
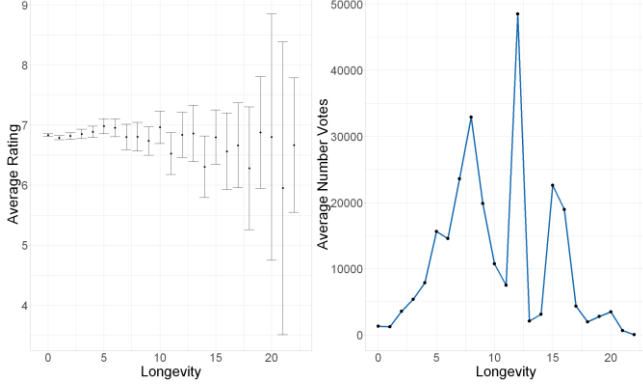

Fig. 4. TV Series Longevity relation with (a) Average Rating, (b) Number of Votes

As shown in Figure 4, no clear trend exists between the longevity and the already defined success quantifiers. Figure 4 (a) makes it clear that with the increasing age the series only become more conflicted, since the average rating fluctuates around the same values, but the variance increases considerably. Figure 4 (b) looks to align with the expectation, increasing the number of votes for series with up to 10 years, but afterwards the trend becomes negative reaching a minimum past 20 years.

Important to note that the spike in the average number of votes when the series reaches 12 years of age is a consequence of the most iconic and beloved shows of this generation all sharing that longevity, as shown in Table 1. These 20 shows together not only amount to 7740276 votes but also average 8.9 rating, making 12 years the clear peak of longevity.

TABLE I. TOP 20 TV SERIES

Title	Average Rating	Number of Votes
Game of Thrones	9.2	2060874
Breaking Bad	9.5	1837827
Stranger Things	8.7	1154541
The Walking D.	8.1	972109
Sherlock	9.1	915102
The B.B. Theory	8.2	799803
Chernobyl	9.4	735906
Dexter	8.7	721404
How I met your M.	8.3	677081
The Office	9.0	581034
True Detective	8.9	563677
Lost	8.3	549239
Black Mirror	8.8	537088
Peaky Blinders	8.8	533493
Prison Break	8.3	530213
Vikings	8.5	525761
Better Call Saul	8.9	505816
Rick and Morty	9.2	505162
House of Cards	8.7	501447
Westworld	8.5	499399

Furthermore, as a way to validate what was previously mentioned on longevity but also to verify if there were any other suiter to success quantifiers, a correlation matrix, displayed in Figure 5, was created.
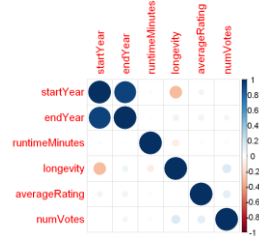

Fig. 5. Potential Success Quantifiers Correlation Matrix

Again, the established conclusions about longevity were reinforced and in addition the start and end years and runtime, as possible quantifiers do not bear any meaningful correlation with the solid established qualifiers, average rating and number of votes.

With this being said, it becomes apparent that, with the data available, the most adequate success qualifiers are average rating and number of votes. So now a criteria must be created to distinguish the bad and middling shows from the great ones. For that and in order to also guarantee a considerable sample size for the top series, the standard created will follow the Equation 1. Where the top TV series must have an average rating greater than 8.6 and more than 3960 number of votes.

$$\begin{cases} Top_{Avg.\ Rating} > Q_{Avg.\ Rating}(0.75) \\ Top_{Num.\ Votes} > Q_{Num.\ Votes}(0.75) \end{cases} \quad (1)$$

IV. BRINGERS OF SUCCESS

Following the creation of an outline for success and, in turn, a way to reliably measure it, the next step on the analysis is to identify what distinguishes the shows.

From the data available the most prominent differentiators of TV series can be grouped in four groups:

- People: containing everyone present on the cast.
- Genre: categorizing the shows by their family of titles.
- Episode runtime: time of each episode of the series.
- Streaming platform: gives insight on what platform houses the shows.

A. People

From a common sense perspective, it only makes sense that the people responsible for the making of a show would have the most impact when it comes to the title's success. Directors and writers have a strong presence in the minds of the most knowledgeable people within the medium and the marketability of actors has been a growing asset for studios and producers alike.

1) Actors

Known to be the face of the series they are in, as well as being plastered on all the promotional material, leading actors carry a great weight on the show's marketing. It is only natural that their physical attributes and their disposition on camara would change to better fit the target audience's needs, as displayed in Figure 6.

Fig. 6. Age and Sex Distribution over (a) Time, (b) Average Rating

From Figure 6 (a) the most apparent that can be taken is the upwards trend that can be seen up to the 2010's, followed by a steep in 2015. This indicates that the average lead actors age has increased over time, which by be an indication of a bias to repeatedly use established leading actors on shows. When taking sex into account, both actors and actresses follow a similar behaviour, but even though the average age is similar the deviation is considerably higher for women.

Figure 6 (b) carries some important information about how the distribution of the age evolves over the rating. From the plot it's possible to identify that the age is very scattered on ratings up to 6, but the values quickly gather on around 45 years with higher ratings. Additionally, it can also be seen that the age of actresses also stabilizes on lower values when compared to the age of actors.

Regarding the possible use of already experienced leading actors, Figure 7 shows how their usage changes over time.
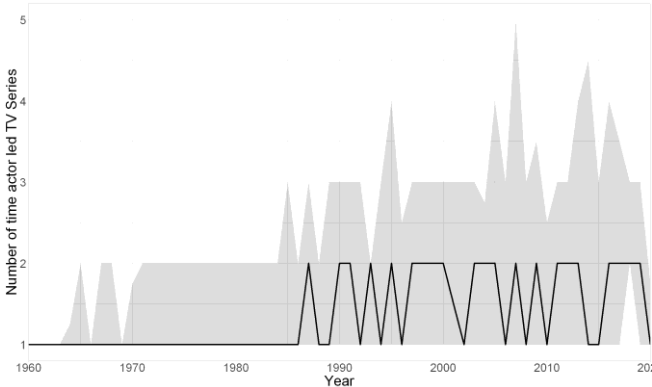

Fig. 7. Evolution of lead actor re-usage over time

With the plot it's possible to see that, especially in more recent years, using repeated leads is favoured, but not to the extent expected from Figure 6 (a). Nevertheless, it becomes important to check if this kind of repeatability is prominent on the highly rated shows or just facilitated by the limited talent in the industry.
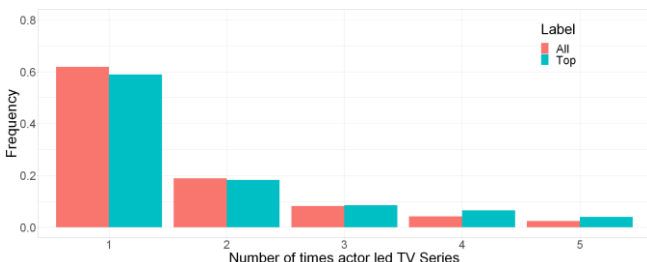

Fig. 8. Frequency of Lead Actors Repeatability

Following Equation 1, Figure 8 displays how all shows compare to the top regarding the leading actor's reuse. For the most part the frequencies are similar, with the top shows slightly benefiting more the repurpose of actors. Regardless, with such miniscule differences is hard to draw any meaningful conclusions.

2) Cast
Even though often overshadowed by the lead actor the remaining cast is often more impactful on the creation of a beloved and compelling story. Staying in line with the previous analysis of the actors, the most reliable way to measure the competence of the cast is their work experience.
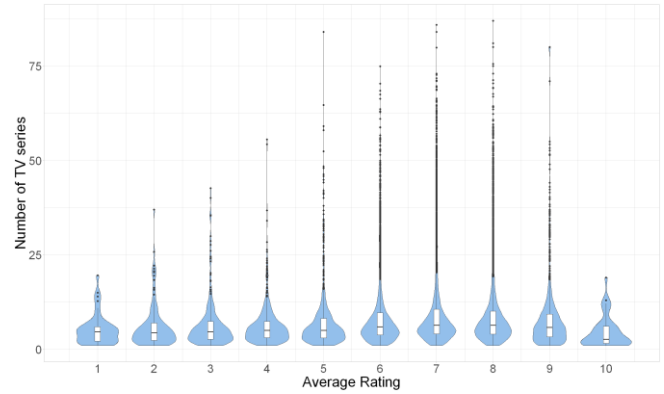

Fig. 9. Cast Experience over Average Rating

Figure 9 shows how the number of shows the cast took part in changes with the average rating, drawing a clear relation between the experience of the cast and the success of the show. Interesting to note that even though for the most part the plot follows expectations, since the average rating increases with the increase in cast experience, the same cannot be said for the higher ratings. From 8.5 onward a fall in experience can be seen, reaching the minimum for all ratings at 10. This is probably a consequence of the limited sample size for series with a rating of 9.5 or higher, but nevertheless opens questions on how impactful cast experience truly is.

3) Directors and Writers
Manly staying behind the scenes, it is unequivocal the presence and importance the writers and directors have on the making of a TV series. Even though not as tangible for the average viewer as their actor counterparts, the orientation they provide can make or break a show. With the dataset available, director information is scarce and unreliable therefore the writers will be focused on the analysis.
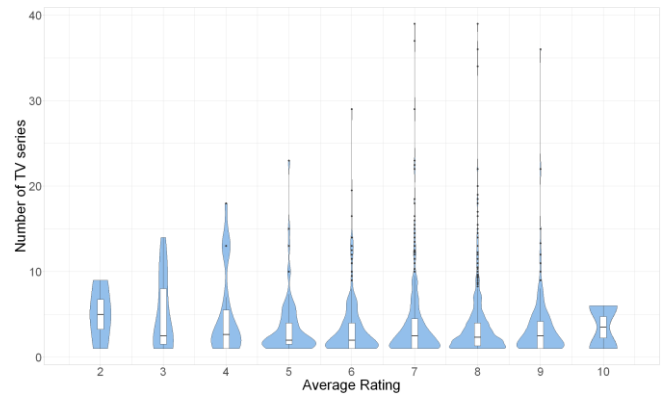

Fig. 10. Writers Experience over Average Rating

Figure 10 breaks expectations, showing that most when making the series writers lack any prior experience and the ones that do lack any meaningful relation with the shows rating. The relationship between the values lacks any trends and establishes no correlation. This is emphasised by the two ratings with highest writer average experience are in opposites of the scale, being 2 and 10.

B. Genres

When it comes to differentiating shows between each other genre is arguably the best resource. Being a categorical metric with distinct easy-to-understand characteristics it widely used as a descriptor for TV series. Figure 11 displays what genres are favoured by the industry as well as which ones garner the most attention by fans.
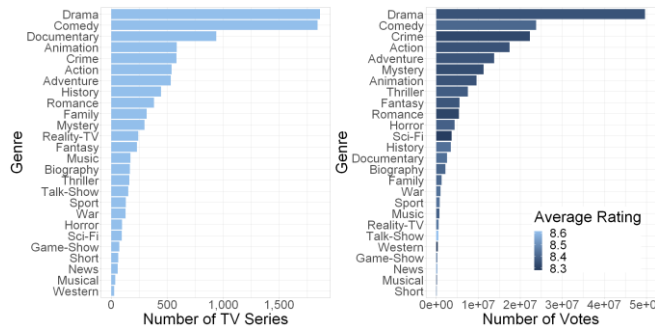

Fig. 11. Genres over (a) Number TV Series, (b) Number of Votes

From the plots it is possible to compare the amount of investment given to each genre with an overall appreciation of the viewers. Even though with small variations, the displacement of the genres is quite consistent between the two figures, which brings forward the question, "is the number of votes a consequence of the number of movies or is the production of movies a reflection of the viewer feedback?". When trying to answer this relation, Figures 12 (a) and 12 (b) respectively display, over time, the number of movies made, and average rating for each of the 3 most produced genres.
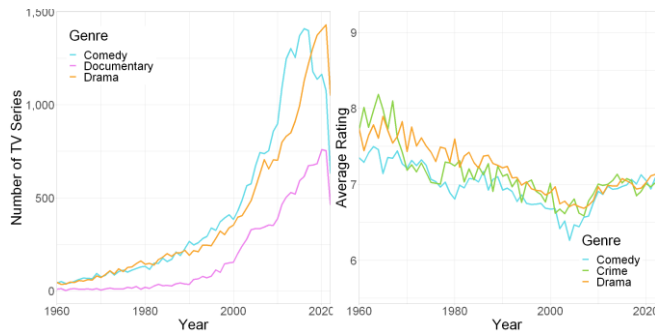

Fig. 12. Top 3 genres over time (a) Evolution of the number of movies produced for the most frequent genres, (b) Evolution of average ratings for the movies with most viewer feedback (number of votes)

The number of votes increases exponentially through the years, matching the increase in TV series production. The considerable differences seen for each genre also prove that some are preferred over other and this inclination changes in a dynamically with the times. Other than that, as expected from Figure 11 (b) the average rating follows a remarkably similar pattern for every genre, so Figure 13 was created to better display its distribution.
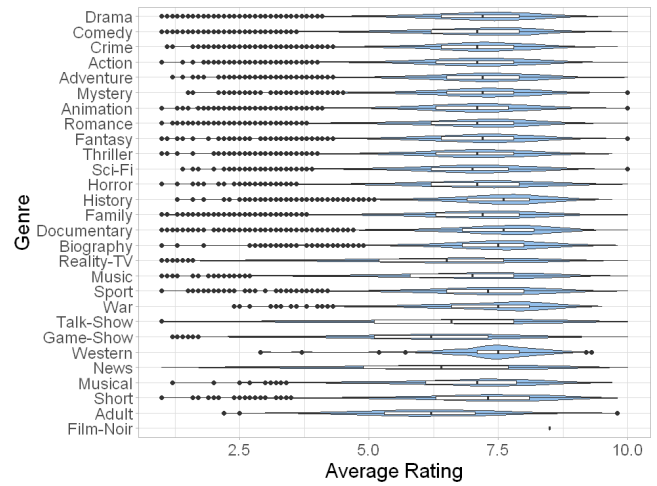

Fig. 13. Average Rating distribution for each Genre

The differences in mean and standard deviation shown on the plot strengthens the idea that some genres are preferred by the public, while others do not carry the same appeal.
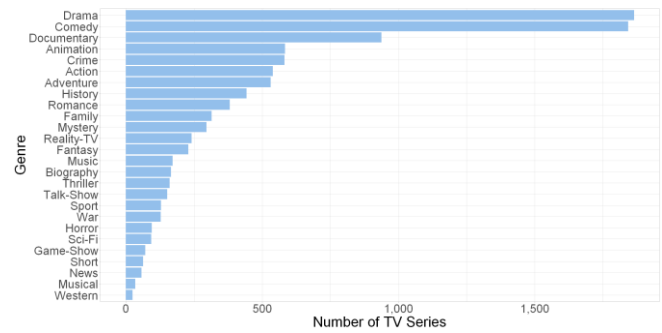

Fig. 14. Number of Top TV Series by Genre

Figure 14 shares the genres of the top TV series, identified by Equation 1. The plot closely resembles Figure 11 (a) with some small differences such as the filtering of the least popular genres, Film-Noir and Adult, and probably the most impactful the demotion of Reality-TV from the top 5 genres.

C. Runtime

Time is often treated as the most valuable currency a person can have. In the constant battle the entertainment industry faces to grab people's attention and make the best out of their time, the runtime of a series is certainly something to consider. Figure 15 displays how the duration is distributed after filtering for outliers and unreasonable faulty durations present on the dataset.
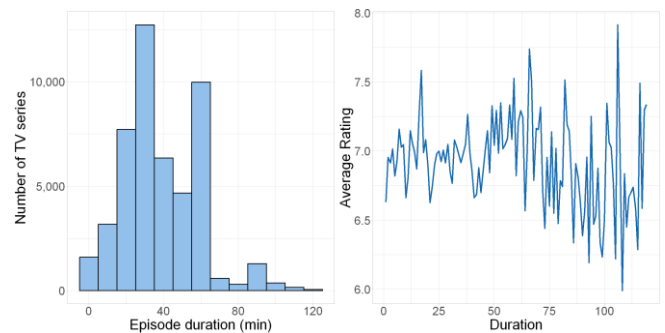

Fig. 15. Episode duration distribution (a) Histogram, (b) Evolution of average rating over episode duration

From Figure 15 (a) it is possible to see that most TV series have episodes runtime between 30 and 40 minutes. This together with the fact that around 90% of the show's duration fall under the 70 minutes, clearly indicates a preference for lower times. Figure 15 (b) also displays similar information, with a downwards trend started following the 70-minute mark. It can also be observed that the longer the runtime, the more unpredictable the ratings are as there doesn't seem to be a clear relationship between rating and runtime for episodes above 70 minutes in duration.

When evaluating the same distribution this time for the top shows the same behaviour can be analysed, as seen in Figure 16. The bias for low runtimes is still present with the 70-minute threshold still being synonymous with a sharp drop off in production. Additionally, TV series within 60 to 70 minutes are now favoured encompassing most of the top shows.
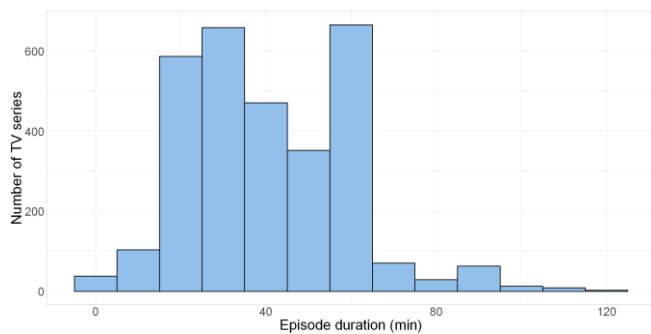

Fig. 16. Top series episode duration distribution

D. Streaming Platform

The growing presence of streaming services on the entertainment industry make them an essential resource when trying to build a new successful show. Figure 17 shows how the TV series are distributed within the 4 most prominent streaming platforms.
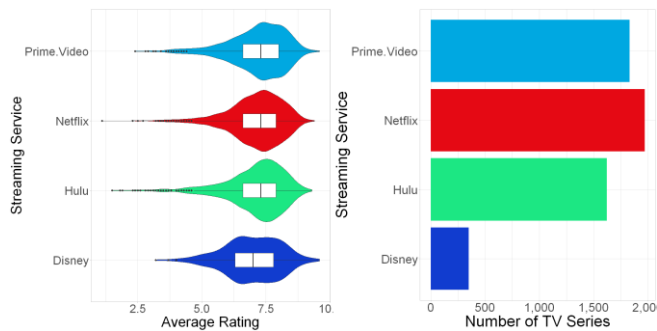

Fig. 17. Streaming platforms information (a) Distribution of rating for each service, (b) Number of TV series under the platform

Netflix being the pioneer of the format still leads the pack when it comes to total number of shows available to watch. That being said both Prime Video and Hulu are catching up their filmography while maintaining an overall better average rating. Disney on the other hand is still very recent and has fallen behind the other on both volume and quality of shows. While Netflix, Prime Video and Hulu have a mean average of about 6.75, Disney+ has an average closer to 6.5. However,

this difference only accounts for 2.5% so it shouldn't be relevant.

Netflix and Hulu have a lot of outliers on the lower ratings, which, when considered alongside the number of productions each streaming service has, indicates that Prime Video is the best choice for quality and quantity in productions out of these options.

V. CONCLUSION

Success is an extremely relative concept that can be interpreted in thousands of ways. This work intrinsically linked its definition to the viewer by establishing rating as the main criteria of success. By focussing on TV series, it allowed for a more concise analysis where the unique characteristics of the shows can be better specified, and a more tailored solution can be created.

It should be mentioned that the findings of this study constitute correlations, not necessarily causations, which means that these recommendations are simply rules of thumb that one can follow to increase their chances of success.

Regarding leading actors differences in sex are quite meaningless on their own, only showing more considerable differences taking age into account, where actress are preferred to be younger than their male counterparts. The actor's repeatability shows some promise considering that on average series tend to reuse actors once. In respect to age the relation becomes a lot clearer, since the shows with higher rating tend to have leading actor with age ranging from 40 to 50 years old.

The cast doesn't show a direct correlation with success. That said, a more experience cast is often preferred because it is usually aligned with higher rating.

Genre is the most apparent indicator that some series are bound to be less success on a conceptual level. Clear differences between viewer feedback and the genres. From the data available Comedy and Drama shows dwarf the remaining genres from a production standpoint as well as engagement from the public, being the two most appetizing for producers.

Runtime also brings forward important information on how the community in large enjoys watching shows. Series with more than 70 minutes an episode are scarce and usually diminish the viewers rating. That said, the most wanted shows fall either on the 30 to 40 or 60 to 70 minute ranges. Regarding the platform where the show will be aired, finding the ideal fit is quite difficult. Excluding Disney, the remaining tree both provide almost equal filmographies while maintaining a consist level of quality.

In conclusion, from the analysis made the conditions necessary to maximize the probability of success would boil down to a Drama or Comedy with a lead with age ranging from 40 to 50 years, a jointed cast experience of 50 shows or more, an episode runtime averaging 60 to 70 minutes and finally being streamed on Prime Video.

REFERENCES

IMDb: Ratings, Reviews, and Where to Watch the Best Movies & TV Shows. (n.d.). IMDb. https://www.imdb.com/

TV shows on Netflix, Prime Video, Hulu and Disney+ (Version 3). (2021). [Dataset]. https://www.kaggle.com/datasets/ruchi798/tv-shows-on-netflix-prime-video-hulu-and-disney